Regulation of Mammalian Nucleotide Metabolism and Biosynthesis

Andrew N. Lane  
*University of Kentucky*, andrew.lane@uky.edu

Teresa W-M Fan  
*University of Kentucky*, teresa.fan@uky.edu

Follow this and additional works at: [https://uknowledge.uky.edu/toxicology_facpub](https://uknowledge.uky.edu/toxicology_facpub)

もあります Part of the Medical Toxicology Commons

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Repository Citation

Lane, Andrew N. and Fan, Teresa W-M, "Regulation of Mammalian Nucleotide Metabolism and Biosynthesis" (2015). *Toxicology and Cancer Biology Faculty Publications*. 34.  
[https://uknowledge.uky.edu/toxicology_facpub/34](https://uknowledge.uky.edu/toxicology_facpub/34)

This Article is brought to you for free and open access by the Toxicology and Cancer Biology at UKnowledge. It has been accepted for inclusion in Toxicology and Cancer Biology Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
Regulation of Mammalian Nucleotide Metabolism and Biosynthesis

Digital Object Identifier (DOI)
http://dx.doi.org/10.1093/nar/gkv047

Notes/Citation Information
Published in Nucleic Acids Research, v. 43, issue 4, p. 2466-2485.

© The Author(s) 2015. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
REGULATION OF MAMMALIAN NUCLEOTIDE METABOLISM AND BIOSYNTHESIS

Andrew N. Lane* and Teresa W.-M. Fan

Graduate Center of Toxicology and Markey Cancer Center, University of Kentucky, Biopharm Complex, 789 S. Limestone St, Lexington, KY 40536, USA

Received September 21, 2014; Revised December 21, 2014; Accepted January 12, 2015

ABSTRACT

Nucleotides are required for a wide variety of biological processes and are constantly synthesized de novo in all cells. When cells proliferate, increased nucleotide synthesis is necessary for DNA replication and for RNA production to support protein synthesis at different stages of the cell cycle, during which these events are regulated at multiple levels. Therefore the synthesis of the precursor nucleotides is also strongly regulated at multiple levels. Nucleotide synthesis is an energy intensive process that uses multiple metabolic pathways across different cell compartments and several sources of carbon and nitrogen. The processes are regulated at the transcription level by a set of master transcription factors but also at the enzyme level by allosteric regulation and feedback inhibition. Here we review the cellular demands of nucleotide biosynthesis, their metabolic pathways and mechanisms of regulation during the cell cycle. The use of stable isotope tracers for delineating the biosynthetic routes of the multiple intersecting pathways and how these are quantitatively controlled under different conditions is also highlighted. Moreover, the importance of nucleotide synthesis for cell viability is discussed and how this may lead to potential new approaches to drug development in diseases such as cancer.

INTRODUCTION

A large fraction of the genome is now known to be transcribed into a wide range of RNAs whose functions are still being ascertained (1–4). Even in quiescent cells, there is considerable turnover of RNA involved in cell maintenance, repair and regulation. Proliferating cells must up-regulate RNA and DNA biosynthesis as an essential component of cell division, which can modulate, at least in part, the rate of the overall cell cycle (5,6). This requires increased expression of the genes associated with nucleotide synthesis in late G1 phase (5–15).

Nucleotide synthesis is regulated by several critical transcription factors (cf. Table 1), MYC and Rb/E2F in particular, which if mutated or overexpressed are associated with transformation and uncontrolled proliferation leading to cancer (16–23). MYC directly regulates the expression of genes that encode the enzymes in the nucleotide biosynthetic pathways and in the feeder pathways for the production of the precursors of all nucleotides (15,24–26), as well as coordinates RNA and protein biosynthesis (27,28). MYC also influences expression of specific microRNAs that regulate enzymes required for cell proliferation (22,29–31).

Despite its functional importance, nucleotide metabolism and cell cycle control have received much less attention than genomics and functional genomics, although there have been multiple metabolic targets derived from the relevant processes for human disease therapy, such as the antimetabolites MTX (32,33), gemcitabine (34), purine analogues (35), suicide inhibitors like 5-FU (36), a range of antiviral nucleotide analogues (37,38) and traditional RNA-seeking antibiotics (39).

Although numerous recent reviews have dealt with metabolic adaptations in proliferating cells (40–50), there has been little emphasis on nucleotide biosynthesis and its regulation. Here we review the regulation of energy and metabolic pathways needed for nucleotide biosynthesis in proliferating mammalian cells.

CELLULAR CONTENT OF NUCLEOTIDES AND NUCLEIC ACIDS

RNA and DNA content of mammalian cells

The DNA content of cells in an organism is fixed, and does not depend on the cell size. In contrast, other cellular com-

*To whom correspondence should be addressed. Tel: +859 2182868; Email: andrew.lane@uky.edu

© The Author(s) 2015. Published by Oxford University Press on behalf of Nucleic Acids Research. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Table 1. Genes and enzymes involved in nucleotide biosynthesis

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Enzyme</th>
<th>Chromosome location</th>
<th>Reaction (Figure No.)</th>
<th>Predicted and known regulators</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Purines</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRPS</td>
<td>Phosphoribosyl Pyrophosphate Synthetase</td>
<td>Xq22.3</td>
<td>(Supplementary Figure S1)</td>
<td>GR Sox5, p53 FOXD1 Nkx2-5 FOXO3b POU2F1 POU2F2a FOXO4 GR-alpha Bach1, GATA-1, Bach2, C/EBPa, CHOP-10, Brachyury, Roa, Arnt, USF-1; MYC</td>
</tr>
<tr>
<td>PPAT</td>
<td>phosphoribosyl pyrophosphate amidotransferase</td>
<td>4q12</td>
<td>1 (3)</td>
<td>Bach1, GATA-1, C/EBPa, CHOP-10, Brachyury, Roa, Arnt, USF-1; MYC</td>
</tr>
<tr>
<td>GART</td>
<td>Trifunctional enzyme. Phosphoribosylglycinamide Formyltransferase Phosphoribosylglycinamide Synthetase Phosphoribosylaminoimidazole Synthetase; Phosphoribosylaminomimidazole Synthetase</td>
<td>21q22.11</td>
<td>2.3.5 (3)</td>
<td>SRF, HoxA9, HoxA9b, Meis-1, CUTL1, Stat5a, Fox O3/3a/3b, MYC</td>
</tr>
<tr>
<td>PFAS</td>
<td>Formylglycinamide ribotide amidotransferase</td>
<td>17p31.1</td>
<td>4 (3)</td>
<td>AML1a, MAX, MAX1, MYC</td>
</tr>
<tr>
<td>PAICS</td>
<td>Bifunctional Phosphoribosylaminomimidazole Carboxylase</td>
<td>4q12</td>
<td>6, 7 (3)</td>
<td>AREB6, p53, C2, STAT3, MyoD, MYC</td>
</tr>
<tr>
<td>ADSL</td>
<td>Adenylosuccinate Lyase</td>
<td>2q35</td>
<td>9, 10 (3)</td>
<td>FOXO1a, C/EBPb, SREBP-1a, b, c, NF-κB, AP-1, c-jun; MYC</td>
</tr>
<tr>
<td>ATIC</td>
<td>5-Aminomimidazole-4-Carboxamide Ribonucleotide Cyclodehydrogenase</td>
<td>2q35</td>
<td>9, 10 (3)</td>
<td>FOXO1a, C/EBPb, SREBP-1a, b, c, NF-κB, AP-1, c-jun; MYC</td>
</tr>
<tr>
<td>IMPDH1</td>
<td>IMP dehydrogenase</td>
<td>7q31.3-q32</td>
<td>13 (3)</td>
<td>c-Fos, USF1, AP-1, NRF-2, USF-1, c-Jun; MYC</td>
</tr>
<tr>
<td>GMPS</td>
<td>GMP synthetase</td>
<td>3q25.31</td>
<td>13 (3)</td>
<td>E2F-4, E2F-3a, E2F-5, E2F-1, E2F p53</td>
</tr>
<tr>
<td>ADSS</td>
<td>Adenylosuccinate synthetase</td>
<td>1q44</td>
<td>11 (3)</td>
<td>USF1, Pax-5, NRF-2, USF-1, c-Jun, MYC</td>
</tr>
<tr>
<td>ADSL</td>
<td>Adenylosuccinate Lyase</td>
<td>22q13.2</td>
<td>12 (S2)</td>
<td>Bach1, GATA-1, Bach2, C/EBPa, CHOP-10, Brachyury, Roa, Arnt, USF-1, MYC</td>
</tr>
<tr>
<td>AK1</td>
<td>Adenylate kinase</td>
<td>9q34.1</td>
<td>8 (3)</td>
<td>MYC</td>
</tr>
<tr>
<td>NME</td>
<td>Nucleoside Diphosphate Kinase</td>
<td>17q21.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2. Pyrimidines</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAD, trifunctional</td>
<td>Carbamoyl-Phosphate Synthetase 2</td>
<td>2p22-p21</td>
<td>1 (Figure 2)</td>
<td>PPAR-gamma1 AP-1, ATF-2, MYD</td>
</tr>
<tr>
<td>DHOHH</td>
<td>Aspartate Transcarbamylase</td>
<td>16q22</td>
<td>4 (Figure 2)</td>
<td>AdR, AML1 a p300, CUTL1, NF-kappaB</td>
</tr>
<tr>
<td>UMPH</td>
<td>Bifunctional Uridine Monophosphate Synthetase: orotate phosphoribosyltransferase</td>
<td>3q13</td>
<td>5 (Figure 2)</td>
<td>POU2F1a, ER-alpha, AML1a, NF-κB</td>
</tr>
<tr>
<td>UMPS</td>
<td>Monophosphate Synthetase: orotate phosphoribosyltransferase</td>
<td>3q21.2</td>
<td>6 (Figure 2)</td>
<td>POU2F1a, ER-alpha, AML1a, NF-κB</td>
</tr>
<tr>
<td>NME</td>
<td>Uridine decarboxylase</td>
<td>17q21.3</td>
<td>13 (3)</td>
<td>FOXO1b, TF, POU2F1a, POU2F1b, MYC</td>
</tr>
<tr>
<td>CTPS</td>
<td>CTP Synthase 1, 2</td>
<td>1p34.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>3. ribose/PPP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G6PD</td>
<td>Glucose-6-phosphate dehydrogenase</td>
<td>Xq28</td>
<td>(Figure 1)</td>
<td>TBP, p53, ATF-2, c-Jun</td>
</tr>
<tr>
<td>H6PD</td>
<td>D-glucono-1,5-lactone lactone hydrolase (also G6PD activity)</td>
<td>1p36</td>
<td>(Figure 1)</td>
<td>ER-alpha, Sp1, p53, G6PD</td>
</tr>
<tr>
<td>PGD</td>
<td>6-phosphogluconate dehydrogenase (deficiency not associated with disease)</td>
<td>1p36.22</td>
<td>(Figure 1)</td>
<td>RhoA, NF-1, c-Myc</td>
</tr>
<tr>
<td>RPIA</td>
<td>ribose 5-phosphate isomerase A</td>
<td>2p11.2</td>
<td>(Figure 1)</td>
<td>MYC, GR, AML1a, NCX, NCx, MYD</td>
</tr>
<tr>
<td>RPE</td>
<td>ribulose 5-phosphate epimerase</td>
<td>2q32-q33.3</td>
<td>(Figure 1)</td>
<td>Sp1, AP-1, ATF-2, SRF</td>
</tr>
<tr>
<td>TALDO1</td>
<td>Transaldolase 1</td>
<td>11p15.5-p15.4</td>
<td>(Figure 1)</td>
<td>SRF, POU2F1, CUTL1, HNF-3beta, YY1, FOXO1, FOXO2</td>
</tr>
<tr>
<td>TKT</td>
<td>Transketolase</td>
<td>3p14.3</td>
<td>(Figure 1)</td>
<td>NF-1, Sp1, HNF-1, LUN-1, Egr-4, C/EBPalpha</td>
</tr>
<tr>
<td><strong>4. Feeder pathways</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GOT1</td>
<td>Glutamic-Oxaloacetic Transaminase (cytoplasmic)</td>
<td>10q24.1-q25.1</td>
<td>(Figure 1)</td>
<td>STAT1, STAT3, Stat1alpha</td>
</tr>
<tr>
<td>GOT2</td>
<td>Glutamic-Oxaloacetic Transaminase (mitochondrial)</td>
<td>16q21</td>
<td>(Figure 1)</td>
<td>AREB6, Egr-2, PPAR-gamma1, FOXO1a</td>
</tr>
<tr>
<td>GLUD1</td>
<td>Glutamate dehydrogenase 1</td>
<td>10q23.3</td>
<td>(Figure 1)</td>
<td>PPAR-gamma2, FOXO1, MYC, MIIF-1, NF-kappaB, GATA-2, AREB6, SRY, POU2F1a, FKBP12, NOF</td>
</tr>
</tbody>
</table>
components depend on cell volume, as the concentration is regulated. Thus, the protein concentration of mammalian cells is about 200 mg/ml (20% solution), which can occupy ∼16% of the cell volume, not counting the shells of 'bound' water in macromolecules. However, cell volumes vary widely - by more than an order of magnitude even for a given organism (see below). This means that the macromolecular content per cell with the exception of DNA varies over a factor of 10-fold or more from one cell type to another.

Quiescent mammalian cells in G0 or G1 are typically diploid, and contain the minimum amount of both DNA and RNA. In order to pass into S phase, the genes for DNA biosynthesis must first be upregulated. Furthermore, actively proliferating cells must double other macromolecular content as they enter M phase and divide into two daughter cells. As the major macromolecular component of cells is protein, protein biosynthesis must also be greatly upregulated during S phase. This will require an increase in the number of ribosomes and thus rRNA, as well as the concomitant energy production needed to meet the enhanced demand for the highly endergonic nucleotide biosynthesis (7,13,51). However, it has been argued that the utilization for macromolecule biosynthesis is a small fraction of the total cellular adenosine triphosphate (ATP) consumption (52,53). Nevertheless, the ATP concentrations (and probably the more relevant ATP/ADP:Pi ratio which is the thermodynamic potential (54,55)) cannot fall below critical levels to maintain cell viability. The ATP requirements show a hierarchy of proliferation > maintenance and repair > membrane potential (46,56–57), i.e. proliferation ceases first, while maintaining membrane potentials goes last. However in mouse Ehrlich Ascites cells, no such energy hierarchy was detected (58).

The content of mammalian cell biomass has been documented for different conditions (59,60). However, it is important to specify the cell type and conditions, whether it be cells in G0/G1 or proliferating cells, as well as the cell size. For example, all proliferating cells double their biomass as they complete a cell cycle. However, in the diploid state, the amount of DNA per cell is independent of the cell size, as it is fixed by the genome C-value (61). For human diploid cells, this is about 6.5 pg per cell based on 6 billion base pairs. However, cell volumes vary greatly, at least by an order of magnitude from small hematopoietic cells of 0.2–0.5 pl to large hepatocytes of volume 5 pl (60,62–63). As other macromolecules are maintained at about the same concentration per cell, the biomass content scales with cell size, as has been emphasized recently by Vaquez et al. (51). Biomass accounts for around 30% of the cell volume (i.e. 70% water v/v). Given the relative abundances of the various macromolecules, their effective partial specific volumes and the density of mammalian cells, (64–66), (67) the biomass content of a cell is ca. 0.35–0.4 g/ml. In resting cells, biomass comprises ∼50–60% protein, 15% lipid, 5% carbohydrates, 5% small metabolites and ions, and the remainder the nucleic acids (i.e. ca. 15–20%). In a small cell, the DNA content may be up to 3% of the biomass, but in a large cell it is a much smaller fraction (cf. 0.2%). The remaining nucleic acid content is RNA, of which 85% is rRNA and <5% is mRNA. Mammalian cells have been said to have an RNA content of 3% of the biomass (68). However, this seems to be rather variable depending on cell type and size (58). Note that this is considerably less than the >20% estimated for Escherichia coli (52).

The RNA content of a cell does not necessarily represent the amount of RNA synthesis however, because the RNAs

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Enzyme</th>
<th>Chromosome location</th>
<th>Reaction (Figure No.)</th>
<th>Predicted and known regulators</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSAT1</td>
<td>Phosphoserine Aminotransferase 1</td>
<td>9q21.2</td>
<td>3 (Figure 4)</td>
<td>POU2F2 (Oct-2.1) oct-B3 oct-B2 POU2F2 POU2F1 c-Jun POU2F2B AREB8 GR CREB p53 deltaCREB SEF-1 (1) GR-alpha Nkx2-5</td>
</tr>
<tr>
<td>PSPH</td>
<td>Phosphoserine Phosphatase</td>
<td>7p11.2</td>
<td>4 (Figure 4)</td>
<td>AHR CHOP-10-CBF-B CBF-A NF-YA c-Myc PIAS NF-Y CBF(2) EBPalpha</td>
</tr>
<tr>
<td>SHMT1</td>
<td>Serine Hydroxymethyltransferase 1 (cyto)</td>
<td>17p11.2</td>
<td>5 (Figure 4)</td>
<td>E2F-3a E2F-1 Spl E2F-2 GATA-1</td>
</tr>
<tr>
<td>SHMT2</td>
<td>Serine Hydroxymethyltransferase 2 (mito)</td>
<td>12q12-q14</td>
<td>5 (Figure 4)</td>
<td></td>
</tr>
<tr>
<td>TYMS</td>
<td>Thymidylate synthetase</td>
<td>18p11.32</td>
<td></td>
<td>AML1a p53 ATF-2 Egr-1 RRRE-1 FAC1 POU2F1 POU2F1a ARP-1 MRF-2 MYC USF1 Spl p53 USF-1 USF-2 EBPalpha Pax-3 POU2F1 POU2F1a USF-1</td>
</tr>
<tr>
<td>DHRF</td>
<td>Dihydrofolate reductase</td>
<td>5q11.2-13.2</td>
<td></td>
<td>RFX1 NF-YA NF-YC STAT5A CBF-C NF-YB HoxA5 POU2F1 POU2F1a POU2F2</td>
</tr>
<tr>
<td>MTHFD1</td>
<td>Trifunctional: Methylenetetrahydrofolate Dehydrogenase (NADP+ Dependent) 1, Methylenetetrahydrofolate Cyclohydrolase, Formyltetrahydrofolate Synthetase</td>
<td>14q24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RRM1</td>
<td>Ribonucleotide reductase subunit 1</td>
<td>11p15.5</td>
<td></td>
<td>E2F-4 E2F-3a E2F-5 E2F-2 POU2F1 HoxA5 Lmo2 E2F-2, MYC CREB AP-2alpha isofrom 3 Spl AP-1 deltaCREB AP-2alpha isofrom 4 AP-2alpha isofrom 2 AP-2alpha AP-2alpha</td>
</tr>
<tr>
<td>RRM2</td>
<td>Ribonucleotide reductase subunit 2</td>
<td>2p25-p24</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The gene nomenclature and regulation (including allosteric regulators) refer to mammalian systems. Predicted transcription factors are from http://www.genecards.org/, which recognize consensus binding sites in the promoters of given genes. Factors in bold represent those experimentally verified by at least one method. MYC binding to gene promoters is mainly from Liu (15); MYC regulation of TS, IMPDH1, PRPS2 is from (38) and E2F from (17). MYC regulation of nucleotide biosynthesis was confirmed by Kam et al. (39).
are synthesized as larger precursors that are then processed to the mature size, with recycling of the released mononucleotides. For example the ratio of intron to exon length in the human genome is about 28 (69) and the length of mature rRNAs are about 50% of the pre rRNA transcript (70). This requires at least three-fold more RNA synthesis than the amount of mature RNA present in the cell.

Increased protein biosynthetic rate scales with cell size and progression into S-phase. To make protein faster, the cell must make more ribosomes, which requires more rRNA and therefore an increase in the rate of production of rNTPs, which suggests that RNA and protein biosynthesis are coordinately regulated. It has been shown that MYC and its downstream target translation initiation factor eiF4E achieve this via controlling the cis-regulatory element in the 5’ UTR of phosphoribosylpyrophosphate synthetase (PRPS2) (27), which catalyzes the first committed step in nucleotide biosynthesis (cf. Figures 1 and 3A).

**Cellular concentrations of rNTPs and ATP requirement for nucleic acid metabolism**

To sustain energy and precursors for RNA biosynthesis, the ATP concentration in cells is generally maintained above 1 mM (71) with a high ATP/ADP ratio (72), so that there is sufficient free energy available to drive endergonic reactions. There are limits to the maintenance of these levels, below which different aspects of the cell functions cease, eventually leading to necrotic cell death (57,73–74). Some cells maintain very high concentrations of ATP (>5 mM) for specialized mechanical work, such as skeletal muscle (75) and cardiac myocytes (76). Similarly, other rNTPs are maintained at sufficient concentrations to activate and drive anabolic processes (see above) while supplying the materials for nucleic acid synthesis. Thus in cells, it is typically observed that the nucleotide concentrations are in the order ATP > UTP > GTP > CTP with the CTP level kept at ~1 μmol/g (77,78). Nucleotide sugars are also needed to activate metabolites for various anabolic processes including UDP-glucose and other UDP-hexoses (for carbohydrates synthesis), CDP-choline (for lipid synthesis), GDP sugars (e.g. GDP mannose for glycosyltransferases), NAD(P)⁺, FAD/FMN (for mediating redox reactions) and ADP ribosylation for a wide range of regulatory functions (79,80). Collectively these compounds are typically present at >1 mM. For a cell of 1 pl volume, the amount of nucleotides is of the order 4–10 fmol. To maintain these pools at constant concentrations as a cell divides, these nucleotides must be synthesized.

For a cell to divide, the entire biomass including metabolites (which is cell size dependent) is doubled as it progresses from the beginning of G1 through the cell cycle and cytokinesis. As described above, a cell with a volume of 0.2–5 pl has a biomass content of 0.08–2 ng, comprising roughly 0.05–1.2 ng protein, 0.01–0.3 ng RNA, 6.5 pg DNA, 0.01–0.3 ng lipid and 0.004–0.1 ng each of carbohydrates and metabolites. The nucleic acids for cell doubling alone amount to 50–950 fmol nucleotides, which use eight ATP equivalents on average for de novo synthesis or 0.4–8 pmol ATP. However, as the RNAs are synthesized as larger precursor molecules and energy is required for recycling the released nucleoside monophosphate (NMPs), the actual ATP usage for RNA synthesis must be larger. For nucleic acid biosynthesis, which is energetically costly, the nucleotide synthesis consumes 0.5–9 pmol exogenous carbon, which is comparable in number to the ATP hydrolysis. Thus, to make nucleic acids for cell proliferation purposes, cells have to upregulate both energy metabolism and the nucleotide biosynthetic pathways. As expected nucleotide biosynthesis is greatly stimulated as cells enter rapid growth (7,11,16,81). Progression through the cell cycle is tightly regulated by numerous transcription factors, and is associated with changes in volume, energy and anabolic metabolism as cells progress through S-phase (15–16,49,51,62,81–87).

**NUCLEIC ACID SYNTHESIS: ENERGETICS AND NUTRIENT REQUIREMENTS**

To maintain homeostasis, dividing cells need to replenish nucleotides at the same rate as cell division. Thus, the progression of the cell cycle must be tightly linked to the ability of the cell to acquire nutrients, generate metabolic energy and to drive anabolism, including nucleotide/nucleic acid biosynthesis. Although there are salvage pathways and cells can take up nucleotides (88–90), most proliferating cells synthesize nucleotides and nucleic acids de novo, mainly from glucose, glutamine and CO₂. The metabolic demands of nucleic acid synthesis have been reviewed recently (47).

**Bioenergetics of nucleotide biosynthesis**

The different parts of the nucleotides derive from various carbon and nitrogen sources in the cell (cf. Figures 1–3 and below), and the assembly of the mature rNTPs has a high metabolic demand. Supplementary Table S1 summarizes the numbers of nucleoside triphosphate (NTP) equivalents (number of phosphates released from ATP and GTP) needed to make one molecule each of the four rNTPs, according to Figures 1–3. Thus starting from glucose, three ATP equivalents are needed to make the activated ribose-5'-phosphoribosylpyrophosphate (PRPP), which is produced by the reaction of 5'-phosphoribose with ATP, driven by the release of the good leaving group 5'-AMP (Supplementary Figure S1). The pyrimidine rings are synthesized first as uracil from aspartate, CO₂ (or bicarbonate) and glutamine (Figure 2), which requires two ATP. One of the biosynthetic steps, i.e. orotate dehydrogenase reaction, occurs in the mitochondria (91), while the remainder reside in the cytoplasm. The amination of U to C consumes an additional ATP; de novo synthesis of 5'-UMP and 5'-CMP therefore require four and five molecules of ATP respectively. Aspartate provides three of the four carbon atoms of pyrimidines, which derive largely from glutamine and to a lesser extent from glucose (92) (and see below).

Unlike pyrimidines, the synthesis of purine nucleotides is entirely cytoplasmic, with the nucleobase being built directly on the activated PRPP (again, using three ATPs starting from glucose) (Figure 3A). The five carbon atoms of the purine ring derive from CO₂, glycine and the one-carbon unit N¹⁰-formyl-TetraHydroFolate (THF), which is derived from the serine-glycine pathway via N⁵,N¹⁰-methylenetetrahydro-THF (Figure 4). The sources of serine and glycine may
Nucleic Acids Research, 2015, Vol. 43, No. 4

Figure 1. Denovo nucleotide biosynthesis: generation of activated ribose. 5-phosphoribose-1-pyrophosphate (PRPP) is the activated form of ribose used for nucleotide biosynthesis and is derived from ribose-5-phosphate from the pentose phosphate pathway (PPP). Ribose-5-phosphate is produced via both oxidative and non-oxidative branches of the PPP. The oxidative branch also generates two NADPH. The oxidative branch comprises the reactions catalyzed by G6PD, PGLS and PGD. The non-oxidative branch interconverts five carbon sugars with four and six carbon sugars using the transaldolase (TA) and transketolase (TK) reactions. HK: hexokinase; G6PD: glucose-6-phosphate dehydrogenase; PGLS: 6-phosphogluconolactonase; PGD: 6-phosphogluconate dehydrogenase; RPI: ribulose-5-phosphate isomerase; RPE: PGLS 3-epimerase; TK: transketolase; TA: transaldolase.

be exogenous, and/or via de novo synthesis from predominantly glucose (93–96). The production of the common intermediate in purine biosynthesis, Inosine Monophosphate (IMP), uses seven ATPs (Figure 3A; Supplementary Table S1). An additional GTP or ATP is used to convert IMP into AMP or GMP, respectively (Figure 3B). The NMPs are then converted to the triphosphate nucleotides by the action of nucleotide kinases, which use two additional ATP molecules for a total of 10. Hence for a genome of 50% GC content and 6 \times 10^{10} nucleotides, DNA replication alone would consume around 6 \times 10^{10} ATPs per cycle or close to 0.1 pmol ATP/cell and around 15 fmol glucose equivalents in terms of carbon. The energy demand for RNA synthesis may be 10-fold higher during a cell cycle, thus accounting for about 1 pmol ATP/cell.

The production of DNA further requires the reduction of rNDPs to dNDPs via ribonucleotide reductase, which is NADPH-dependent (97), therefore requiring the supply of around 0.1 pmol NADPH/cell per division (and see below). The source of this NADPH is variable according to cell type (98), but the oxidative branch of the pentose phosphate pathway (PPP) is an efficient cytoplasmic source, which also links glucose metabolism to ribose production. As in general the DNA content of cells is lower than that of RNA, and compared with rNTPs, dNTPs have fewer functions other than nucleic acid synthesis, the concentrations of the dNTPs relatively low in cells, typically in the micromolar range in cells in G1, and rising ca. 5–10-fold in late G1 or during S phase (99,100).

In addition to NADPH, as alluded to earlier, the nucleotide biosynthesis pathways have feeder pathways that provide for the carbon and nitrogen precursors, including the amino acids aspartate, glutamine, serine and glycine as well as CO2. These feeder pathways are glycolysis, the PPP (Figure 1), the serine-glycine pathways (Figure 4), the Krebs cycle without or with anaplerotic inputs (Figure 5) and glutamine amidotransferase reactions (Figures 2 and 3).

The metabolic energy needed to drive nucleotide biosynthesis is therefore substantial, and is expected to be derived from a combination of glycolysis and oxidative phosphorylation. The ATP yields for oxidation of different substrates assuming perfectly coupled mitochondria are given in Supplementary Table S2. Different cell types use different strategies that are also dependent on the tissue environment including the nutrient supply, which is a point of focus for understanding metabolic reprogramming in cancers (25,44,47,101–107). Although the oxygen concentration in tissues (e.g. solid tumors) can become quite low, reaching <1% or 10 \mu M (108,109), it is still sufficient to saturate cytochrome c oxidase, which has a low \mu M K_m for oxygen (110,111). However, other metabolic reprogramming occurs at oxygen levels lower than about 2–4%, due to the activation of HIF1α (112,113), which leads to accelerated glycolysis but inhibition of the Krebs cycle activity. HIF1α is constitutively degraded by the proteasome under normoxia
Figure 2. Pyrimidine biosynthesis. CA: carbamoyl aspartate; DHO: dihydroorotate; OMP: orotate monophosphate. Enzyme names: (1) carbamoyl phosphate synthase II (CPSII); (2) aspartate transcarbamoylase (ATCase); (3) carbamoyl aspartate dehydratase = dihydroorotase \([CAD\text{ encodes enzymes } 1 + 2 + 3]\); (4) dihydroorotate dehydrogenase; (5) orotate phosphoribosyltransferase; (6) orotidine-5-phosphate decarboxylase (OMP decarboxylase). The activities of 5 and 6 reside in a single bifunctional polypeptide encoded by the \(UMPS\) gene. Atom colors denotes origins: red from \(\text{CO}_2\), green from aspartate and ultimately glucose or Gln, blue from Gln.

However, hypoxic cells also ferment glucose to lactate, producing 2 ATP/mol glucose and in cancer cells glycolysis is typically accelerated many fold \((42,121–122)\), which may produce ATP at a rate comparable to the more efficient mitochondrial oxidation of fuels including fatty acids \((104,123)\), Gln \((25,44–45,81,106,119,124–127)\) and ketone bodies \((46)\) (Supplementary Table S2). It is notable that in the interstitial fluid of solid tumors, the glucose levels are very low with depletion of lipids compared with the blood supply in the tumor \((128)\). This could reflect the high energy demand of tumor cells, which oxidize nutrients at a high rate.

Enzymes and gene locations involved in nucleotide biosynthesis and ‘feeder’ pathways

Figures 1–4 and Table 1 lists the enzymes and genes directly involved in the \textit{de novo} biosynthesis of nucleotides as well as...
the relevant ‘feeder pathways’ that supply the carbon, nitrogen and phosphorus, as well as metabolic energy. The genes are distributed over nine chromosomes for the purines and five for the pyrimidines, plus several others for the ‘feeder pathways’. The expression of the genes is thought to be controlled by several transcription factors, though relatively few have been directly demonstrated experimentally.

A notable feature of the enzymes involved in nucleobase biosynthesis is the multifunctional nature, including PPAT, GART and MTHFD1 for purines and carbamoyl phosphate synthase II (CPSII) aspartate transcarbamoylase (ATCase) dihydroorotase trifunctional enzyme (CAD) and UMPS for pyrimidines. These enzymes are thought to be rate limiting in nucleotide biosynthesis, though in practice this concept may not be helpful in understanding regulation.
Coordinate regulation of expression and flux control

The rate of cell growth is ultimately limited by the supply of materials and energy, and therefore by the flux through the relevant metabolic pathways. Nucleic acid biosynthesis, like any other complex intersecting metabolic networks requires coordination. The flux through even a simple linear pathway depends on the supply of the initial metabolite, the concentration and activities of all of the enzymes in the pathway, and the presence of any feedback or other regulatory controls. It is not well appreciated that the control of flux generally does not reside in any single enzyme or even the same enzymes under all conditions (129–134). Fundamentally, at steady state, the flux is the same at every point in the pathway. The concept of a rate-limiting enzyme is therefore at best conditional. In metabolic control analysis (MCA), the goal is to decipher how much control each enzyme in a pathway exerts on the net flux (which can be non-zero only where the system is not at equilibrium). A rate-limiting step in a multistep process is one that controls the net flux, such that changes in any of the other steps in the process have no influence on the rate. Glycolysis for example, is often considered to have three rate-limiting enzymes, HK, PFK-1 and PK. A single rate limiting enzyme is one that completely controls the pathway flux and its flux coefficient (FCC) is unity. FCC is defined as $\partial \ln J / \partial \ln v_i$, where $J$ is the net flux and $v_i$ is the enzymatic activity of

(129–132). This is because a coordinate activation of multiple enzymes is necessary for enhancing flux through the pathway (131,133).
the \( i \)th enzyme. If there are three equally rate limiting steps, then the flux control is shared equally among the three enzymes. As the sum of the FCC is unity, then each enzyme now has an FCC of one-third (131). In the context of MCA, this is equivalent to having the FCC of 0.33 for each of these enzymes, which is a form of distributed flux control. This means that changing the concentration of any one of these enzymes by 10-fold will change the flux by only a factor of 2, compared with a 10-fold change in flux if these enzymes were simultaneously changed 10-fold. In poorly vascularized tumors, however, the flux control resides mainly at the glucose uptake and HK steps (46,135).

Distributed control is a general feature of metabolic pathways. Even if there is one enzyme that is rate limiting, increasing its activity indefinitely will only transfer the rate limiting step to other enzymes in the pathway, thereby limiting the effect of up regulating a given enzyme activity (by expression or other enzyme-level activity control).

In pyrimidine biosynthesis, the rate limiting step has been described as CPSII (the first enzyme of CAD) (11) and dihydroorotate dehydrogenase (DHODH) (136), the fourth step in the pathway (Figure 2), again implying distributed flux control. This and the fact that several enzymes in both purine and pyrimidine biosynthesis have activities residing on a single polypeptide chain indicates a level of coordinate, stoichiometric expression and the possibility of channeling intermediates from one enzyme to another (see next section).

A means to coordinate expression of functionally linked enzymes that are coded for on different chromosomes is to make use of a common or a set of transcription factors. Although the promoters of the nucleotide genes have binding sites for a variety of transcription factors, there are many in common, one that stands out is MYC (along with related factors that interact with MYC (e.g. MAX and E2F), which have been independently verified by validation experiments such as MYC promoter occupancy assays (Table 1). The regulation of the expression of many genes by MYC is also linked to signaling pathways that respond to growth sensing cues, such as the EGFR/MAP kinase pathway (5,137–139), Hif1/2α (140–142) or estrogen receptor/Sp1 (137,143) for CAD regulation. Although not yet completely understood, cross-talk among these gene regulatory networks is likely to add additional complexities to homeostatic control.

In addition to transcriptional-level regulation, direct modulation of various enzyme activities of the pathways is important, as described below.

**Substrate level control**

Several of the enzymes involved in nucleotide biosynthesis are regulated at the enzyme level by allosteric interactions, primarily by feedback inhibition. For example, the activities of the enzymes PPAT and CAD multifunctional enzymes are strongly inhibited by relevant pathway products (11,139). Purine biosynthesis is inhibited by AMP and GMP and Pi that act on PRPP synthetase (Supplementary Figure S1) (7) and by ATP and GXP (adenosine and guanosine mono, di or triphosphates) at two sites on the PRPP amidotransferase (Figure 3) (144–146). It is also evident that PRPP amidotransferase activity may be stimulated by the substrate PRPP (147). The branch point at IMP to AMP and GMP (Supplementary Figure S2) is further regulated by allosteric interactions, in which AMP and GMP inhibit the adenylosuccinate synthetase and IMP dehydrogenases, respectively (148–150). The enzyme activities therefore depend not only on the concentrations of the immediate substrates, but also on the end products. Substrate availability in eukaryotic cells can be an important means of controlling flux through pathways, either by physical compartmentation (important in pyrimidine biosynthesis) or by diffusion limits between different pools in large cells (151).

A potentially powerful means of overcome kinetic/diffusion limitations is to channel intermediates in multienzyme complexes (152,153). The observation that several of the enzymes in mammals are expressed as multifunctional enzymes (Figures 2 and 3, Table 1), and that these activities do not always catalyze sequential reactions has lead to the hypothesis that there might be a purinosome (91), in which intermediates are channeled rather than freely equilibrating with the bulk cytoplasm. Whereas the evidence for specific stoichiometric associations in vivo is weak, kinetic experiments in vitro are consistent with channeling, implying the possibility of weak and transient complex formation (154). Putative purinosomes bodies have been detected in vivo (91), but they may also be artifacts of the constructs used for visualization (155).

The pyrimidine synthesis pathway is controlled by feedback inhibition by UTP acting on the CPSII domain of the CAD trifunctional enzyme (Figure 2, Table 1) (5,11), which is also activated by PRPP, thereby integrating regulation of purine and pyrimidine biosynthesis (11). As for purine biosynthesis, mammalian pyrimidine biosynthesis is characterized by multifunctional enzymes (Table 1), in contrast to the homologues of *E. coli*, where CPS and ATCase are two non-associated polypeptides and are differentially regulated at the metabolite level (11). For efficient capturing of the CAD product dihydroorotate (DHO) by the mitochondrial DHODH, and the observation of the association of both CAD and mitochondria with the cytoskeleton, it has been suggested that CAD could be translocated via the cytoskeleton to the DHODH site (11). One complication is that up to 30% of the CAD may translocate to the nucleus during the S phase of the cell cycle, when demand for pyrimidine biosynthesis is maximal. This would compromise the DHODH efficiency and favor an alternative moonlighting function for nuclear CAD (11). However, DHO could still be readily diffuse from the nucleus into the mitochondria as the latter are in close proximity to the nuclear membrane (156).

**Synthesis of the deoxyNTPs**

DNA synthesis requires a source of the four dNTPs, which derive from the ribonucleotides at the level of rNDPs by reduction at the 2′ position of the ribose subunit. The synthesis of dNTPs occurs cytoplasmically using the enzyme ribonucleotide reductase (RNR) (157), though the possibility of RNR activity inside mitochondria has been reported (158). This enzyme acts upon the rNTP and reduces the 2′-OH of ribose to the deoxy state, using NADPH as the
electron the incorporation of the number and type of atoms (isotopologues) into nucleotides. NMR in contrast depends on the difference in magnetic properties of the nuclei of $^{13}\text{C}$ versus $^{12}\text{C}$ for example. $^{12}\text{C}$ is magnetically silent, whereas $^{13}\text{C}$ can either be directly detected or indirectly by its influence on the directly attached proton (92,169–170) which has the advantage of greatly enhanced sensitivity. NMR detects both the position (isotopomers) and the amount of incorporation of the labeled atoms. The analysis of the isotopomers and isotopologues of intermediates and products provides very detailed information about the biosynthetic routes to the products under study (25,167). Thus use of $^{13}\text{C}$ and/or $^{15}\text{N}$ labeled precursors coupled with labeled product analysis by MS and NMR offers the optimal approach for this purpose, as these techniques can directly detect the number and positions of $^{13}\text{C}$ or $^{15}\text{N}$ in the intermediates and products of nucleotide biosynthesis, thereby robustly defining the flow of atoms through the intersecting pathways. Other stable isotopes can also be used to measure nucleotide synthesis and incorporation into RNA and DNA. Hellerstein’s group (14,171) has used both $^{13}\text{C}$ glucose and $^2\text{H}$ glucose with GC-MS detection of urine nucleotide released from DNA to follow DNA synthesis in cells and tissues. The overall turnover of nucleic acids in tissues can also be determined by measuring the incorporation of deuterons from $\text{D}_2\text{O}$ into DNA (172). Recently $^{14}\text{C}$-1 and $^{14}\text{C}$-6 glucose was used in comparison with $[1^2\text{H}]$-glucose or $[3^2\text{H}]$-glucose with LC MS to detect the transfer of a deuterons to NAD$^+$ in the oxidative branch of the PPP (166).

Our group has developed the stable isotope resolved metabolomics or SIRM approach (173–175) for this purpose, which has been applied to mapping nucleotide biosynthesis in cultured cells or tissues in situ (19,92,101,168,173,176–177). Figure 5 tracks the expected incorporation of labeled atoms from $^{13}\text{C}_6$-glucose ($\bullet$ for $^{13}\text{C}$) and $^{13}\text{C}_5^{15}\text{N}_2$-glutamine ($\bullet$ for $^{13}\text{C}$ and $\bullet$ for $^{15}\text{N}$) into nucleotide products via purge (PUR) and pyrimidine (PYR) synthesis as well as feeder pathways including glycolysis, the Krebs cycle in the absence or presence of anaplerotic pyruvate carboxylation or glutaminolysis, PPP and one-carbon pathway. Based on SIRM profiling of the various labeled intermediates and products, these intersecting pathways for nucleotide biosynthesis can be rigorously reconstructed and quantified even in human subjects in situ (154). Examples of stable isotope tracer studies on nucleotide biosynthesis are described below.

Tracking ribose synthesis via the pentose phosphate pathway

The ribose unit of the nucleotides derives from the PPP, and for most cells and tissues, this means that as few tissues are gluconeogenic the carbon originates from glucose either exogenous or via G-1P by phosphorolysis of glycogen. Two-dimensional (2D) NMR can be applied to crude cell or tissue extracts and discriminates the ribose subunit of purine and pyrimidine nucleotides as well as those of NAD$^+$. The TOCSY experiment detects the scalar interactions between the ribose H1′–H2′–H3′–H4′ which appear as discrete cross peaks. When the carbon atom is $^{13}\text{C}$, the cross peaks are single, whereas if $^{15}\text{C}$ is present, it splits the attached proton res-
onance into two peaks separated by the one-bond coupling constant of 130–160 Hz. The patterns of the so-called satellite peaks are characteristic according to which atoms are enriched in $^{13}$C, which gives rise to the isotopomer distribution that can be quantified (167,178). Complementary isotope edited experiments can be used to determine the complete isotopomer distribution in complex systems as needed (177,179). This is straightforward to demonstrate in the free nucleotide pool in cells by 2D $^1$H NMR analysis, as shown in Supplementary Figure S3. Here the ribose subunits of the free nucleotide pools are highly enriched with $^{13}$C when the cells are exposed to $^{13}$C glucose, but not at all when exposed to $^{13}$C Gln. The production of $^{13}$C$_5$-ribose-containing nucleotides from $^{13}$C$_6$-glucose, as determined by FT-ICR-MS analysis (5 $^{13}$C isotopeologue in Supplementary Figure S4) is also consistent with the PPP activity. The use of glucose as the primary carbon source for ribose biosynthesis applies to many different cell types (92,168,175,179–180).

However, it is difficult to delineate whether the $^{13}$C$_5$-ribose labeling pattern results from oxidative, non-oxidative or both branches of the PPP. The oxidative branch (Figure 1) generates ribose-5-phosphate from glucose-6-phosphate (G6P) via decarboxylation of the C1 of G6P while producing two molecules of NADPH. This reaction is especially important in cells that have a high demand for NADPH, including erythrocytes (to remove H$_2$O$_2$ in a highly oxidizing environment (181)) and proliferating cells that are actively making fatty acids (166). The alternative non-oxidative branch reversibly converts ribose-5-phosphate and erythrose-4-phosphate into fructose-6-phosphate and glyceraldehyde-3-phosphate, and vice versa (182), which serves a dual purpose, namely to direct glucose carbon back into glycolysis and for nucleotide ribose biosynthesis. This branch does not generate NADPH. Boros et al. have established a simple GC-MS based stable isotope tracing method that can discriminate between these two branches of pathways, while estimating the net lactic fermentation flux (12,183–186). This method utilizes $^{13}$C$_2$-1,2-glucose as tracer, which undergoes decarboxylation to become exclusively $^{13}$C-1-ribose via the oxidative branch of the PPP, whereas via extensive scrambling in the non-oxidative branch, ribose is labeled at the C1 and C2 positions. This scrambling arises from the reversibility of the reactions catalyzed by transaldolase and transketolase (cf. Figure 1A), which in fact may be exacerbated by the exchanges in the individual half reactions (187). Chemical exchange in reactions that are near equilibrium equilibrate label across the reaction, such that there can be label reaching a metabolite even where there is net flux in the opposite direction. This is a very general problem, that must be accounted for in any atom-resolved tracer experiments. The ratios of the different labeled ribose species provides an estimate of the relative flow through the oxidative and non-oxidative pathways subject to the caveats noted by Kleijn et al. (187). $^{14}$C-1,6 glucose has also been used to discriminate between the oxidative and non-oxidative branches, which relies on radiometric detection of CO$_2$ released by the oxidation at C1’ (166). NMR analysis of free nucleotides can determine positional enrichment in the ribose subunits of intact nucleotides (92) without the need for fragmentation, which may provide a more direct discrimination of between the oxidative and non-oxidative branches and the relative flux through the two branches (188) (and A.N. Lane, unpublished).

It appears that both pathways may operate leading to net nucleotide ribose biosynthesis. Depending on the cell type, and possibly the growth conditions, the oxidative branch can account for very little to most of the flow of glucose into ribose (12,101,182–183,185–186,189–192). Such direct readout of the two pathway activities is required to determine the functional consequences of relevant gene or even enzyme expression in the two pathways (Table 1), as post-translational and substrate level regulations can further alter the metabolic outcome. Despite its complexity and importance to cell proliferation and maintenance, the flux through the PPP is only 1–7% of the net glucose flux to pyruvate (183,188,193–194).

**Tracking pyrimidine ring biosynthesis**

The pyrimidine ring is synthesized independent of ribose-5-P, with all but one reaction occurring in the cytoplasm. The one exception is the DHODH reaction, which resides on the mitochondrial membrane and is intimately coupled to the electron transport chain (ETC). Mitochondria with a defective ETC cannot make UMP. As Figures 2 and 5 show, four of the ring atoms of uracil derived from aspartate, one from CO$_2$ and the N3 from Gln via the cytoplasmic CPSII enzyme. Aspartate is a non-essential amino acid present in blood plasma at about 20 μM (195) and can be transported into the cell by Na$^+$-dependent anionic transporters (196) or synthesized de novo by transamination of oxalacetate (OAA) in either the mitochondria or the cytoplasm. This reaction in fact is an important component of the malate/aspartate shuttle for transferring electrons from NADH in the cytoplasm into mitochondria. In addition to be the entry point of the Krebs cycle, OAA is the product of the anaplerotic reaction catalyzed by pyruvate carboxylase (PC), which is important in some cancers (179,197–198).

If aspartate and thus OAA is used for pyrimidine biosynthesis, anaplerosis is needed to replenish OAA to sustain the functioning of the Krebs cycle. In addition to PC, another common anaplerotic reaction in cancer cells and other proliferating cells (175,194) is glutaminolysis, in which glutamine is first hydrolyzed to glutamate and ammonia followed by either transamination with OAA to produce Asp + α-ketoglutarate (αKG) (199) or oxidative deamination via GDH to αKG + ammonium ions. αKG can also be converted to OAA by the normal functioning of the Krebs cycle (86,125–127,200–204) (Figure 5). The relative importance of these two anaplerotic reactions probably depends on cell type and growth conditions to maintain energy production, nitrogen balance and anabolic metabolism (175,179,194,198,205). The main difference between PC and glutaminolytic anaplerotic pathways is the production of ammonium ions by the latter, which in excess is toxic to cells and must be dealt with. Moreover, glutaminolysis competes with other Gln deamination reactions as the direct nitrogen source for many anabolic processes including nucleotide synthesis (45,206–207). These considerations would probably in part govern the choice of anaplerotic
pathways that proliferating cells adopt to meet their growth demand.

Again, tracer technologies can readily discriminate between glutamine and glucose as carbon sources and the different feeder pathways leading to uracil ring biosynthesis. [U-13C]-glucose enters the Krebs cycle via pyruvate either as 13C2-acetyl CoA (via PDH) or 13C3-OAA (via PC), which in turn gives rise to distinct labeling patterns in citrate, Asp and ultimately the uracil ring. For example, PC- (●) or PDH-derived 13C labeled Asp (●) leads to the synthesis of 13C4, 5, 6-UMP or 13C5, 6-UMP, respectively, after one Krebs cycle turn, as shown in Figure 5. Although [U-13C]-Gln also produces 13C4, 5, 6-UMP (●) via the first Krebs cycle turn, the 13C labeling patterns of Glu, αKG and citrate derived from [U-13C]-Gln are distinct from those derived from [U-13C]-glucose (Figure 5). Quantitative analysis of the isotope distribution in these intermediates and products provides a means of determining the relative contribution of Gln and glucose to uracil biosynthesis (92). It is generally found that in proliferating cells, whether transformed or primary cells that are stimulated to proliferate, Gln is the preferred carbon source for pyrimidine ring over glucose (92), but the precise balance is cell-dependent (Lane & Fan, unpublished data), and possibly also condition dependent.

Cells in culture synthesize a large fraction of the aspartate pool that is used for pyrimidine biosynthesis from glutamine and glucose (81, 92, 101, 179, 208), even in media such as RPMI that contain aspartate (0.15 mM) at concentrations much higher than that present in blood (ca. 20 μM). Also, in cells treated with 13C5, 15N2-Gln, it is found by high resolution FT-ICR-MS that the aspartate pool is enriched mainly in the 13C4, 14N and 13C4, 14N isotopologues, despite the presence of high levels of exogenous aspartate (168). Both labeled Asp species are produced as a consequence of glutaminolysis and transamination. 13C5, 15N2 glutamine produces 13C5, 15N1 glutamate and then 13C5-αKG via glutaminase and aminotransferase activity. αKG enters into the Krebs cycle to produce 13C4-OAA, which is transaminated to form 13C4, 15N1-Asp or 13C4, 14N1-Asp in the mitochondria (Figure 5). The aspartate-malate shuttle then transfers the labeled Asp to the cytoplasm to supply both carbon (C4–C6) and nitrogen (N1) for pyrimidine biosynthesis. Thus, tracking the 13C and/or 15N fate from labeled precursors to the various intermediates and pyrimidine products enable a complete reconstruction of the feeder and pyrimidine synthesis pathways, including the delineation of different compartmental events.

However, there are few examples of this type of analysis in the literature, further studies are needed to determine whether other significant carbon sources are used by some cells (e.g. threonine in stem cells (209)) or under different growth conditions.

Tracking purine ring biosynthesis

The de novo synthesis of the purine ring is considerably more complex than that of the pyrimidine ring, as the former is built up atom by atom on the phosphorylated ribose unit in the cytoplasm (Figure 3). The carbon sources are CO2, glycine and N10-formyltetrahydrofolate (Figure 4). Overall, glycine contributes up to four of the five carbon atoms in the purine ring, including indirectly via N10-formyl THF. Glycine is a non-essential amino acid that is present in human blood at ca. 0.25 mM (210) (HMDB: http://www.hmdb.ca/). It can also be synthesized from glucose via glycolytically produced 3-phosphoglycerate (3-PGA) (Figure 4), via the choline-betaine pathway (211) or from other sources such as threonine in some organisms, though not humans (212). It appears that the glycine precursor for purine biosynthesis is synthesized de novo from serine via glycolysis or mitochondrially via the glycine cleavage system (213) (Figure 4), and there may even be a net efflux of glycine in proliferating cells (214, 215). In fact, exogenous serine, or endogenously synthesized serine, but not exogenous glycine is the major source of the one carbon units for purine biosynthesis (216). The nitrogen in purines derives from glycine (N7), glutamine (N3, N9) and Asp (N1) (Figure 3).

Incorporation of ring nitrogen atoms (N3 and N9) from the amido nitrogen of glutamine is readily observed in the free nucleotide pool using 15N-enriched Glc tracer coupled with 2D 3H(15N)-HSQC NMR analysis. The N3 and N9 of AXP and GXP can be detected in crude cell or tissue extracts using the 2-bond scalar coupling of N9 to C8H or N3 to C2H (cf. AXP data in Figure 6). This 15N incorporation pattern is evident in different cell types (19, 92, 177). In addition, using high resolution FT-ICR-MS, the incorporation of 13C and 15N from [U-13C, 15N]-Gln into nucleotide can be simultaneously tracked by utilizing the difference in the effective neutron mass in 13C versus 15C. This capability makes it practical to reconstruct both carbon and nitrogen pathways without the need for separate experiments (168). Supplementary Figure S4 illustrates FT-ICR-MS analysis of the 13C labeling pattern of AMP derived from [U-13C]-glucose (168). Together with the NMR analysis for 13C labeled positions, we can ascertain that 13C5-AMP represents AMP with fully 13C labeled ribose subunit, while 13C5,9-AMP have fully 13C labeled ribose plus 1–4 13C labels in the adenine ring (and see also (92, 168, 177). These data indicate that PPP, glycolysis, Ser-Gly-N10-formyl THF variably contribute to de novo synthesis of AMP. Thus, NMR, and MS are synergistic in defining the source and detailed labeling patterns for nucleotides, from which the feeder and nucleotide synthesis pathways can be reconstructed and their modulation by transcription factors or environmental conditions can be elucidated.

Recently a tracer approach for following both NAD(P)H production from deuteriated substrates was used to determine the synthesis of glycine in cancer cells and it was shown that the serine/glycine pathway was a net producer of both mitochondrial glycine and NADPH (213, 217).

Salvage pathways

RNA in particular is constantly turned over in cells, both during the production of mature RNAs from longer precursors and to regulate the amounts of, e.g. mRNA. The breakdown of polymeric RNA and DNA results in release of NMPs, which can be recycled by the action of the nucleotide kinases. This process contributes to the nucleotide pools but cannot be used for net synthesis. Cells may also
Nucleic Acids Research, 2015, Vol. 43, No. 4

Figure 6. $^{15}$N incorporation from [U-$^{13}$C,$^{15}$N]-glutamine into purines detected by HSQC. A549 cells were grown the presence of [U-$^{13}$C,$^{15}$N]-glutamine for 24 h. $^{1}H$-$^{15}$N HSQC NMR spectra were recorded at 18.8 T using an INEPT delay optimized for two bond couplings in aromatic systems. The two ring $^{15}$N atoms derived from the amido N of Gln (blue) are indirectly detected by their coupled protons (red) as cross-peaks of N3 to H2 and N9 to H8 in the adenine ring of AXP. Reproduced with kind permission from Springer Science+Business Media from J. Biomolec NMR (Springer) J. Biomol NMR. 2011 April; 49(3–4): 267–280. doi:10.1007/s10858-011-9484-6 (figure 9).

transport nucleobases from the external environment and add the appropriate sugar.

Addition of the nucleobase to ribose is achieved either via the PRPP step for purines, or catalyzed by specific pyrimidine phosphorylase to add the base to ribose-1-phosphate (218). Cytidine and deoxycytidine can be salvaged by the agency of cytidine deaminase, producing uridine or deoxyuridine, and then feed into the uracil pathway (219). This in essence randomizes the UTP/CTP pool and must be accounted for where isotope labeling experiments are performed. Cytidine can also be salvaged more directly via uridine-cytidine kinase producing (d)CMP (220). The relative importance of salvage versus de novo synthesis likely depend on the growth conditions and on the specific tissue. Recently it has been argued that salvage pathways are very important in colorectal cancers for example, albeit without the assistance of isotope tracers (221).

Tracking nucleotide incorporation into nucleic acids.

The free nucleotide pools in cells are straightforward to measure as they are relatively abundant, and are readily labeled using $^{13}$C glucose or $^{13}$C,$^{15}$N glutamine (92,168,177,179). It is also practical to extract both DNA and RNA from $^{13}$C and/or $^{15}$N labeled cells and determine incorporation of exogenously supplied labeled tracers into nucleic acids (via free nucleotides) after hydrolysis to the NMPs (183,185). Although not all of the free nucleotide pool enters nucleic acids, it appears that there is a major common pool of labeled free nucleotides that is used for RNA synthesis (92).

Tracking nucleotide synthesis invivo

The amount of de novo nucleotide biosynthesis correlates well with cell proliferation rates. On the other hand, tissues typically have a low mitotic index, even in tumors (which can be highly heterogeneous and may contain a rather low fraction of cancer cells) (222), so that the amount of de novo nucleotide biosynthesis is small. Nevertheless, using [U-$^{13}$C]-glucose tracing in SCID mice, $^{13}$C incorporation into free nucleotide ribose was observed in some tissues (176), as well as in lung tumors resected from human subjects (175,223). The feeder pathways for nucleotide synthesis haves also been traced in vivo in human glioblastomas using [U-$^{13}$C]-glucose as tracer (194,198). $^{13}$C$_2$-glycine was detected by $^{13}$C NMR in [U-$^{13}$C]-glucose labeled glioblastoma tissues, demonstrating synthesis from glucose via the 3PGA-serine pathways (cf. Figure 4). This is despite the presence of 0.25 mM glycine in blood (210). In addition to the free nucleotide pool which is relatively well sampled by good extraction procedures (92,168,224), it is also possible to determine the enrichment in DNA and RNA (12,92,171–172,183,225). With modern instrumentation, the mass iso-
Nucleotides are used in a wide variety of metabolic functions in all cells, as coenzymes, for regulation, for activating substrates, for anabolic purposes as well as providing the subunits of the nucleic acids. Maintenance of nucleotide levels is therefore fundamental to cellular function. Proliferating cells, whether normal division during embryogenesis or in controlled proliferation in the hematopoietic system and stem cell compartments for example, as well as dysregulated cell division in cancer all need to maintain the supply of nucleotides. Unsurprisingly then the supply of nucleotides is strongly regulated and tied to the cell cycle. This is achieved both at the gene expression level by a variety of transcription factors as well as the substrates level regulation of the large number of enzymes needed to synthesize the nucleotides.

Although the basic nucleotide synthesis pathways are known and their energy demand can be estimated, the specific requirement for nutrient precursors/energy and the regulatory networks for modulating nucleotide biosynthesis in dividing cells remain unclear, particularly in terms of dependence on cell type and pathological conditions. This is in part due to the lack of powerful tools for elucidating the actual paths from nutrient precursors through various feeder pathways to de novo synthesized nucleotides. Furthermore, the amount of RNA synthesis greatly exceeds the amount of polymeric RNA present in a cell, owing to the extensive turnover. Characterization of gene and protein expression in these paths can provide useful clues, but such studies need to be validated by functional pathway analysis.

Stable isotope tracers in conjunction with state-of-the-science metabolomics methodologies is especially well suited for such functional studies not only in model cells and animals but also directly in human subjects. The relatively few stable isotope tracer-based studies reported have already uncovered important new insights into nucleotide biosynthesis, such as preference for endogenously synthesized precursors such as glycine and aspartate over those externally supplied, and how resources are reallocated according to the environmental conditions, especially in pathological conditions such as cancers (‘metabolic reprogramming’ (22,44)). Such information cannot be obtained without tracer methods. These considerations point to the need to be able to assess the nucleotide synthesis according to tissue pathology and nature, for which these new technologies can now be applied with some ease (176,194,197–198,226–227). It is notable that many drugs are targeted at nucleotide synthesis at the level of the availability of nucleotides or act as chain terminators (32–39,228–229).

In the future we expect further systems biochemical advances in systems biochemical approaches in which the detailed energy and anabolic pathways are integrated with the ‘feeder’ pathways and the specific gene expression networks that are linked to the cell cycle and thus nucleotide demand for DNA and RNA synthesis. As these are likely to be cell type and environment specific, the modern high throughput ‘omics approaches, such as SIRM described in this review are well suited to define the biochemical details of normal and pathological cell function, which directly informs the optimum modes for therapeutic intervention. As dysregulations of nucleotide metabolism are commonly involved in human disease pathogenesis, such understanding is expected to have important diagnostic and therapeutic benefits in relevant human diseases including cancer and diabetes.

**CONCLUSIONS AND FUTURE DIRECTIONS**

Stable isotope tracers coupled with state-of-the-science NMR and MS analysis for labeling patterns of relevant metabolites provide an unprecedented opportunity for tracking nucleotide biosynthesis in vivo, even in human subjects, under different tissue contexts and pathological conditions. These advances are expected to greatly accelerate our understanding in the regulation of human nucleotide metabolism, the perturbation of which is crucial to the pathogenesis of many human diseases, including cancer.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**FUNDING**

NIH [R01ES022191-01, R01ES022191-01, R21CA133688-02, 1 U24 DK097215-01A1, 1R21CA133688-02, 1 U24 DK097215-01A1, in part]; Carmen L. Buck endowment (to A.N.L.); Edith Gardner endowment (T.W.M.F.). The open access publication charge for this paper has been waived by Oxford University Press - NAR Editorial Board members are entitled to one free paper per year in recognition of their work on behalf of the journal. Conflict of interest statement. None declared.

**REFERENCES**


212. Edgar, A.J. (2005) Mice have a transcribed L-threonine aldolase/GLY1 gene, but the human GLY1 gene is a non-processed pseudogene. *BMC Genomics*, 6, 32–44.


